DOCUMENT RESUME

ED 244 607 IR 011 129

AUTHOR Gevarter, William B.
TITLE An Overview of Computer-Based Natural Language Processing.
INSTITUTION National Aeronautics and Space Administration, Washington, D.C.
REPORT NO NASA-TM-85635
PUB DATE 83
NOTE 52p.; Best copy available.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Artificial Intelligence; *Computational Linguistics; *Computers; Computer Software; Grammar; Information Processing; *Language Processing; *Man Machine Systems; Program Descriptions; Research Projects; Semantics
IDENTIFIERS Database Management Systems; *Natural Language Processing; Parsing

ABSTRACT

Computer-based Natural Language Processing (NLP) is the key to enabling humans and their computer-based creations to interact with machines using natural languages (English, Japanese, German, etc.) rather than formal computer languages. NLP is a major research area in the fields of artificial intelligence and computational linguistics. Commercial natural language interfaces to computers have recently entered the market and the future looks bright for other applications as well. This report presents an overview of: (1) NLP applications; (2) the three basic NLP approaches; (3) various types of grammars; (4) difficulties in semantic processing of phrases and sentences; (5) knowledge representation (KR); (6) types of syntactic parsing; (7) the relationship among semantics, parsing, and understanding; (8) NLP system types, including question answering, natural language interface (NLI), computer aided instruction (CAI), discourse, text understanding, and text generation systems; (9) the functions, approaches, capabilities, and limitations of NLP systems developed for research purposes; (10) current research NLP systems; (11) the approximate price, developer, purpose, and features of some commercial NLP systems; (12) the state of the art of NLP systems; (13) problems and issues related to language use, linguistics, conversation, processor design, database interfaces, and text understanding; (14) research required in these areas; (15) the principal United States and foreign participants in NLP, and the principal U.S. government agencies funding NLP research; (16) future trends and expectations; and (17) further sources of information. A 39-item bibliography and a glossary are also provided. (Author/ESR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED244607

IR

N83-24193

An Overview of Computer-Based Natural Language Processing

National Aeronautics and Space Administration
Washington, D.C.

1983

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
NTIS

IR011129

NASA Technical Memorandum 85635

N83-24193

An Overview of Computer-Based Natural Language Processing

William B. Gevarter
OFFICE OF AERONAUTICS AND SPACE TECHNOLOGY
NASA Headquarters
Washington, D.C. 20546

NASA
National Aeronautics
and Space Administration

Scientific and Technical Information Branch

1983

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

IR0111829*

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-85635		

4. Title and Subtitle	5. Report Date
AN OVERVIEW OF COMPUTER-BASED NATURAL LANGUAGE PROCESSING	April 1983
	6. Performing Organization Code RTC-6
7. Author(s)	8. Performing Organization Report No.
William B. Gevarter	
9. Performing Organization Name and Address	10. Work Unit No.
National Aeronautics and Space Administration Washington, D.C. 20546	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered Technical Memorandum
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code RTC-6

15. Supplementary Notes

16. Abstract

Computer-based Natural Language Processing (NLP) is the key to enabling humans and their computer-based creations to interact with machines in natural language (like English, Japanese, German, etc. in contrast to formal computer languages). The doors that such an achievement can open have made this a major research area in Artificial Intelligence and Computational Linguistics. Commercial natural language interfaces to computers have recently entered the market and the future looks bright for other applications as well.

This report reviews the basic approaches to such systems, the techniques utilized, applications, the state-of-the-art of the technology, issues and research requirements, the major participants, and finally, future trends and expectations.

It is anticipated that this report will prove useful to engineering and research managers, potential users, and others who will be affected by this field as it unfolds.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Artificial Intelligence, Computational Linguistics, Natural Language, Machine Translation, Data Bases, Cognitive Science, Man-Machine Interface, Query Systems, Computers	Unclassified - Unlimited Subject Category 61

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	54	A04

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1983

AN OVERVIEW OF COMPUTER-BASED NATURAL LANGUAGE PROCESSING*

PREFACE

Computer-based Natural Language Processing (NLP) is the key to enabling humans and their computer-based creations to interact with machines in natural language (like English, Japanese, German, etc. in contrast to formal computer languages). The doors that such an achievement can open have made this a major research area in Artificial Intelligence and Computational Linguistics. Commercial natural language interfaces to computers have recently entered the market and the future looks bright for other applications as well.

This report reviews the basic approaches to such systems, the techniques utilized, applications, the state-of-the-art of the technology, issues and research requirements, the major participants, and finally, future trends and expectations.

It is anticipated that this report will prove useful to engineering and research managers, potential users, and others who will be affected by this field as it unfolds.

*This report is part of the NBS/NASA series of overview reports on Artificial Intelligence and Robotics.

Preceding page blank

ACKNOWLEDGEMENTS

I wish to thank the many people and organizations who have contributed to this report, both in providing information, and in reviewing the report and suggesting corrections, modifications and additions. I particularly would like to thank Dr. Brian Phillips of Texas Instruments, Dr. Ralph Grishman of the NRL AI Lab, Dr. Barbara Groz of SRI International, Dr. Dara Dannenberg of Cognitive Systems, Dr. Gary Hendrix of Symantec, and Ms. Vicki Roy of NBS for their thorough review of this report and their many helpful suggestions. However, the responsibility of any remaining errors or inaccuracies must remain with the author.

It is not the intent of the National Bureau of Standards or NASA to recommend or endorse any of the systems, manufacturers, or organizations mentioned in this report, but simply to attempt to provide an overview of the NLP field. However, in a growing field such as NLP, important activities and products may not have been mentioned. Lack of such mention does not in any way imply that they are not also worthwhile. The author would appreciate having any such omissions or oversights called to his attention so that they can be considered for future reports.

TABLE OF CONTENTS

Page

A. Introduction 1
B. Applications 2
C. Approach 3
1. Type A: No World Models 3
a. Key Words or Patterns 3
b. Limited Logic Systems 3
2. Type B: Systems That Use Explicit World Models 3
3. Type C: Systems That Include Information About the Goals and Beliefs in Intelligent Entities 3
D. The Parsing Problem 3
E. Grammar 4
1. Phrase-Structure Grammar—Context Free Grammar 4
2. Transformational Grammar 4
3. Case Grammar 4
4. Semantic Grammars 5
5. Other Grammars 5
F. Semantics and the Cantankerous Aspects of Language 5
1. Multiple Word Senses 5
2. Modifier Attachment 5
3. Noun-Noun Modification 6
4. Pronouns 6
5. Ellipsis and Substitution 6
6. Other Difficulties 6
G. Knowledge Representation 6
1. Procedural Representations 6
2. Declarative Representations 6
a. Logic 6
b. Semantic Networks 6
3. Case Frames 6
4. Conceptual Dependency 6
5. Frame 7
6. Scripts 7
H. Syntactic Parsing 7
1. Template Matching 7
2. Transition Nets 7
3. Other Parsers 8

TABLE OF CONTENTS (continued)

Page

I. **Semantics, Parsing and Understanding** 10
J. **NLP Systems** 11
1. Kinds 11
a. Question Answering Systems 11
b. Natural Language Interfaces (NLI's) 11
c. Computer-Aided Instruction (CAI) 11
d. Discourse 11
e. Text Understanding 11
f. Text Generation 12
g. System Building Tools 12
2. Research NLP Systems 12
a. Interfaces to Computer Programs 12
b. Natural Language Interfaces to Large Data Bases 13
c. Text Understanding 13
d. Text Generation 13
e. Machine Translation 13
f. Current Research NLP Systems 13
3. Commercial Systems 13
K. **State of the Art** 31
L. **Problems and Issues** 31
1. How People Use Language 31
2. Linguistics 31
3. Conversation 31
4. Processor Design 32
5. Data Base Interfaces 33
6. Text Understanding 33
M. **Research Required** 34
1. How People Use Language 34
2. Linguistics 34
3. Conversation 34
4. Processor Design 34
5. Data Base Interfaces 35
6. Text Understanding 35
N. **Principal U.S. Participants in NLP** 35
1. Research and Development 35
2. Principal U.S. Government Agencies Funding NLP Research 36
3. Commercial NLP Systems 36
4. Non-U.S. 37

TABLE OF CONTENTS (continued)

Page

O. Forecast 37
P. Further Sources of Information 38
1. Journals 38
2. Conferences 38
3. Recent Books 39
4. Overviews and Surveys 39
References 41
Glossary 43

LIST OF TABLES

Page

I. Some Applications of Natural Language Processing 2
II. Natural Language Understanding Systems 14
a. SHRDLU 14
b. SOPHIE 15
c. TDUS 16
d. LUNAR 17
e. PLANES/JETS 18
f. ROBOT/INTELLECT 19
g. LADDER 20
h. SAM 21
i. PAM 22
III. Current Research NLP Systems 25
IV. Some Commercial Natural Language Systems 29

LIST OF FIGURES

Page

1. A Transition Network for a Small Subset of English 8
2. Example Noun Phrase Decomposition 9
3. Parse Tree Representation of the Noun Phrase Surface Structure 9

NATURAL LANGUAGE PROCESSING

A. Introduction

One major goal of Artificial Intelligence (AI) research has been to develop the means to interact with machines in natural language (in contrast to a computer language). The interaction may be typed, printed or spoken. The complementary goal has been to understand how humans communicate. The scientific endeavor aimed at achieving these goals has been referred to as computational linguistics*, an effort at the intersection of AI, linguistics, philosophy and psychology.

Human communication in natural language is an activity of the whole intellect. AI researchers, in trying to formalize what is required to properly address natural language, find themselves involved in the long term endeavor of having to come to grips with this whole activity. (Formal linguists tend to restrict themselves to the structure of language.) The current AI approach is to conceptualize language as a knowledge-based system for processing communications and to create computer programs to model that process.

A communication act can serve many purposes, depending on the goals, intentions, and strategies of the communicator. One goal of a communication is to change some aspect of the recipient's mental state. Thus, communication endeavors to add or modify knowledge, change a mood, elicit a response, or establish a new goal for the recipients.

For a computer program to interpret a relatively unrestricted natural language communication, a great deal of knowledge is required. Knowledge is needed of:

—the structure of sentences
—the meaning of words
—the morphology of words
—a model of the beliefs of the sender
—the rules of conversation, and
—an extensive shared body of general information about the world.

This body of knowledge can enable a computer (like a human) to use expectation-driven processing in which knowledge about the usual properties of known objects, concepts, and what typically happens in situations, can be used to understand incomplete or ungrammatical sentences in appropriate contexts.

Thus, Barrow (1979, p. 12) observes:

> In current attempts to handle natural language, the need to use knowledge about the subject matter of the conversation, and not just grammatical niceties, is recognized—it is now believed that reliable translation is not possible without such knowledge. It is essential to find the best interpretation of what is uttered that is consistent with all sources of knowledge—lexical, grammatical, semantic (meaning), topical, and contextual.

*Or more broadly, as Cognitive Science.

Arden (1980, p. 463) adds:

> In writing a program for u :derstanding languages, one is faced with all the problems of artificial intelligence, problems of coping with huge amounts of knowledge, of finding ways to represent and describe complex cognitive structures, as well as finding an appropriate structure in a gigantic space of possibilities. Much of the research in understanding natural languages is aimed at these problems.

As indicated earlier, natural language communication between humans is very dependent upon shared knowledge, models of the world, models of the individuals they are communicating with, and the purposes or goals of the communication. Because the listener has certain expectations based on the context and his (or her) models, it is often the case that only minimal cues are needed in the communication to activate these models and determine the meaning.

The next section, B, briefly outlines applications for natural language processing (NLP) systems. Sections C to I review the technology involved in constructing such systems, with existing NLP systems being summarized in Section J.

The state of the art, problems and issues, research requirements and the principle participants in NLP are covered in Sections K through N. Section O provides a forecast of future developments.

A glossary of terms in NLP is provided at the back of this report. Further sources of information are listed in Section P.

B. Applications

There are many applications for computer-based natural language understanding systems. Some of these are listed in Table I.

TABLE I. Some Applications of Natural Language Processing

Discourse
- Speech Understanding
- Story Understanding

Information Access
- Information Retrieval
- Question Answering Systems
- Computer-Aided Instruction

Information Acquisition or Transformation
- Machine Translation
- Document or Text Understanding
- Automatic Paraphrasing
- Knowledge Compilation
- Knowledge Acquisition

Interaction with Intelligent Programs
- Expert Systems Interfaces
- Decision Support Systems
- Explanation Modules For Computer Actions
- Interactive Interfaces to Computer Programs

Interacting with Machines
- Control of Complex Machines

Language Generation
- Document or Text Generation
- Speech Output
- Writing Aids: e.g., grammar checking

C. Approach

Natural Language Processing (NLP) systems utilize both linguistic knowledge and domain knowledge to interpret the input. As domain knowledge (knowledge about the subject area of the communication) is so important to understanding, it is usual to classify the various systems based on their repre entation and utilization of domain knowledge. On this basis, Hendrix and Sacerdoti (1981) classify systems as Types A, B or C*, with Type A being the simplest, least capable and correspondingly least costly systems.

1. Type A: No World Models

a. Key Words or Patterns

The simplest systems utilize ad hoc data structures to store facts about a limited domain. Input sentences are scanned by the programs for predeclared key words, or patterns, that indicate known objects or relationships. Using this approach, early simple template-based systems, while ignoring the complexities of language, sometimes were able to achieve impressive results. Usually, heuristic empirical rules were used to guide the interpretations.

b. Limited Logic Systems

In limited logic systems, information in their data base was stored in some formal notation, and language mechanisms were utilized to translate the input into the internal form. The internal form chosen was such as to facilitate performing logical inferences on information in the data base.

2. Type B: Systems That Use Explicit World Models

In these systems, knowledge about the domain is explicitly encoded, usually in frame or network representations (discussed in a later section) that allow the system to understand input in terms of context and expectations. Cullinford's work (Schank and Ableson, 1977) on SAM (Script Applier Mechanism) is a good example of this approach.

3. Type C: Systems that Include Information about the Goals and Beliefs of Intelligent Entities

These advanced systems (still in the research stage) attempt to include in their knowledge base information about the beliefs and intentions of the participants in the communication. If the goal of the communication is known, it is much easier to interpret the message. Schank and Abelson's (1977) work on plans and themes reflects this approach.

D. The Parsing Problem

For more complex systems than those based on key words and pattern matching, language knowledge is required to interpret the sentences. The system usually begins by "parsing" the input (processing an input sentence to produce a more useful representation for further analysis). This representation is normally a structural description of the sentence indicating the relationships of the component parts. To address the parsing problem and to interpret the result, the

*Other system classifications are possible, e.g., those based on the range of syntactic coverage.

computational linguistic community has studied syntax, semantics, and pragmatics. Syntax is the study of the structure of phrases and sentences. Semantics is the study of meaning. Pragmatics is the study of the use of language in context.

E. Grammar

Barr and Feigenbaum (1981, p. 229), state, "A grammar of a language is a scheme for specifying the sentences allowed in the language, indicating the syntactic rules for combining words into well-formed phrases and clauses." The following grammars are some of the most important.*

1. Phrase Structure Grammar—Context Free Grammar

Chomsky (see, for example, Winograd, 1983) had a major impact on linguistic research by devising a mathematical approach to language: Chomsky defined a series of grammars based on rules for rewriting sentences into their component parts. He designated these as, 0, 1, 2, or 3, based on the restrictions associated with the rewrite rules, with 3 being the most restrictive.

Type 2—Context-Free (CF) or Phrase Structure Grammar (PSG)—has been one of the most useful in natural-language processing. It has the advantage that all sentence structure derivations can be represented as a tree and practical parsing algorithms exist. Though it is a relatively natural grammar, it is unable to capture all of the sentence constructions found in most natural languages such as English. Gazder (1981) has recently broadened the applicability of CF PSG by adding augmentations to handle situations that do not fit the basic grammar. This generalized Phrase Structure Grammar is now being developed by Hewlett Packard (Gawron et al., 1982).

2. Transformational Grammar

Tennant (1981, p89) observes that "The goal of a language analysis program is recognizing grammatical sentences and representing them in a canonical structure (the underlying structure)." A transformational grammar (Chomsky, 1957) consists of a dictionary, a phrase structure grammar and a set of transformations. In analyzing sentences, using a phrase structure grammar, first a parse tree is produced. This is called the surface structure. The transformational rules are then applied to the parse tree to transform it into a canonical form called the deep (or underlying) structure. As the same thing can be stated in several different ways, there may be many surface structures that translate into a single deep structure.

3. Case Grammar

Case Grammar is a form of Transformational Grammar in which the deep structure is based on cases—semantically relevant syntactic relationships. The central idea is that the deep structure of a simple sentence consists of a verb and one or more noun phrases associated with the verb in a particular relationship. These semantically relevant relationships are called cases. Fillmore (1971) proposed the following cases: Agent, Experiencer, Instrument, Object, Source, Goal, Location, Type and Path.

---

*Charniak and Wilks (1976) provide a good overview of the various approaches.

The cases for each verb form an ordered set referred to as a "case frame." A case frame for the verb "open" would be:

(object (instrument) (agent))

which indicates that open always has an object, but the instrument or agent can be omitted as indicated by their surrounding parentheses. Thus the case frame associated with the verb provides a template which aids in understanding a sentence.

4. Semantic Grammars

In limited domains, to achieve practical systems, it is often useful, instead of using conventional syntactic constituents such as noun phrases, verb phrases and prepositions, to use meaningful semantic components instead. Thus, in place of nouns when dealing with a naval data base, one might use ships, captains, ports and cargos. This approach gives direct access to the semantics of a sentence and substantially simplifies and shortens the processing. Grammars based on this approach are referred to as semantic grammars (see, e.g., Burton, 1976).

5. Other Grammars

A variety of other, but less prominent, grammars have been devised. Still others can be expected to be devised in the future. One example is Montague Grammar (Dowty et al., 1981) which uses a logical functional representation for the grammar and therefore is well suited for the parallel-processing logical approach now being pursued by the Japanese (see Nishida and Doshita, 1982) for their future AI work as embodied in their Fifth Generation Computer research project.

F. Semantics and the Cantankerous Aspects of Language

Semantic processing, as it tries to interpret phrases and sentences, attaches meanings to the words. Unfortunately, English does not make this as simple as looking up the word in the dictionary, but provides many difficulties which require context and other knowledge to resolve.

1. Multiple Word Senses

Syntactic analysis can resolve whether a word is used as a noun or a verb, but further analysis is required to select the sense (meaning) of the noun or verb that is actually used. For example, "fly" used as a noun may be a winged insect, a fancy fishhook, a baseball hit high in the air, or several other interpretations as well. The appropriate sense can be determined by context (e.g., for "fly" the appropriate domain of interest could be extermination, fishing, or sports), or by matching each noun sense with the senses of other words in the sentence. This latter approach was taken by Reiger and Small (1979) using the (still embrionic) technique of "interacting word experts", and by Finin (1980) and McDonald (1982) as the basis for understanding noun compounds.

2. Modifier Attachment

Where to attach a prepositional phrase to the parse tree cannot be determined by syntax alone but requires semantic knowledge. ***Put the plant in the box on the table***, is an example illustrating the difficulties that can be encountered with prepositional phrases.

3. Noun-Noun Modification

Choosing the appropriate relationship when one noun modifies another depends on semantics. For example, for "apple vendor", one's knowledge tends to force the interpretation "vendor of apples" rather than "an apple that is a vendor."

4. Pronouns

Pronouns allow a simplified reference to previously used (or implied) nouns, sets or events. Where feasible, pronoun antecedents are usually identified by reference to the most recent noun phrase having the same pragmatic context as the pronoun.

5. Ellipsis and Substitution

Ellipsis is the phenomenon of not stating explicitly some words in a sentence, but leaving it to the reader or listener to fill them in. Substitution is similar—using a dummy word in place of the omitted words. Employing pragmatics, ellipses and substitutions are usually resolved by matching the incomplete statement to the structures of previous recent sentences—finding the best partial match and then filling in the rest from this matching previous structure.

6. Other Difficulties

In addition to those just mentioned, there are other difficulties, such as anaphoric references, ambiguous noun groups, adjectivals, and incorrect language usage.

G. Knowledge Representation*

As the AI approach to natural language processing is heavily knowledge-based, it is not surprising that a variety of knowledge representation (KR) techniques have found their way into the field. Some of the more important ones are:

1. Procedural Representations—The meanings of words or sentences being expressed as computer programs that reason about their meaning.

2. Declarative Representations

a. Logic—Representation in First Order Predicate Logic, for example.

b. Semantic Networks—Representations of concepts and relationships between concepts as graph structures consisting of nodes and labeled connecting arcs.

3. Case Frames—(covered earlier)

4. Conceptual Dependency—This approach (related to case frames) is an attempt to provide a representation of all actions in terms of a small number of semantic primitives into which input

*More complete presentations on KR can be found in Chapter III of Barr and Feigenbaum (1981), and in Gevarter (1983).

sentences are mapped (see, e.g., Schank and Riesbeck, 1981). The system relies on 11 primitive physical, instrumental and mental ACT's (propel, grasp, speak, attend, P trans, A trans, etc.), plus several other categories or concept types.

5. Frame—A complex data structure for representing a whole situation, complex object or series of events. A frame has slots for objects and relations appropriate to the situation.

6. Scripts—Frame-like data structures for representing stereotyped sequences of events to aid in understanding simple stories.

H. Syntactic Parsing

Parsing assigns structures to sentences. The following types have been developed over the years for NLP (Barr and Feigenbaum, 1981).

1. Template Matching: Most of the early, and some current, NL programs perform parsing by matching their input sentences against a series of stored templates.

2. Transition Nets

Phrase structure grammars can be syntactically decomposed using a set of rewrite rules such as indicated in Figure 1. Observe that a simple sentence can be rewritten as a Noun Phrase and a Verb Phrase as indicated by:

$$S \longrightarrow NP\ VP$$

The noun phrase can be rewritten by the rule

$$NP \longrightarrow (DET)(ADJ^*)N(PP^*)$$

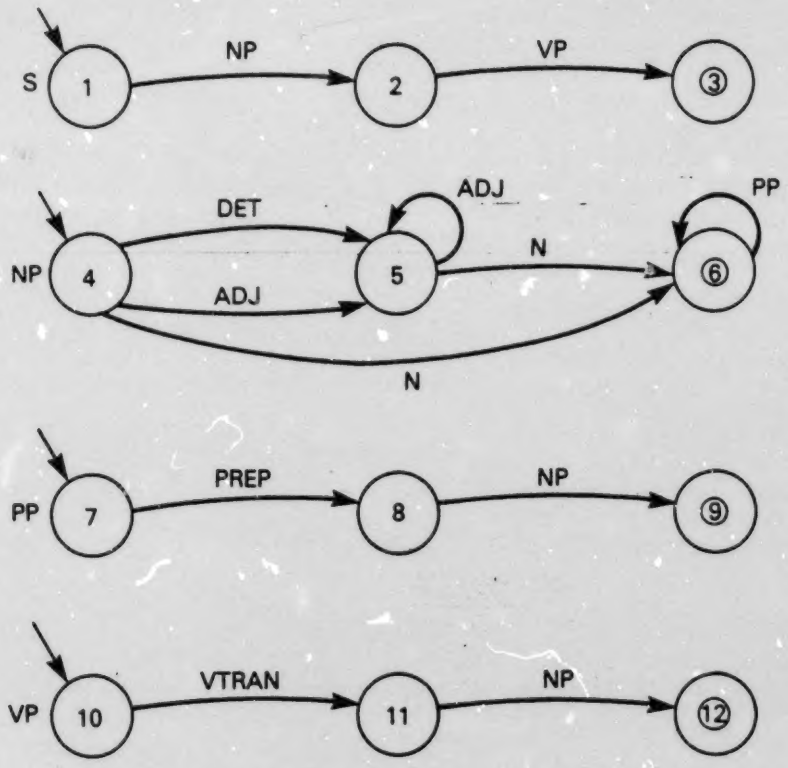
where the parentheses indicate that the item is optional, while the asterisk indicates that any number of the items may occur. The items, if they appear in the sentence, must occur in the order shown. The following example shows how a noun phrase can be analyzed.

$$\underbrace{\text{The large satellite in the sky}}_{NP} \longrightarrow \text{The large satellite } \underbrace{\text{in the sky}}_{PP}$$

-where PP is a prepositional phrase.

Thus, the parser examines the first word to see if it corresponds to its list of determiners (the, a, one, every, etc.). If the first word is found to be a determiner, the parser notes this and proceeds on to the next word, otherwise it checks to see if the first word is an adjective, and so forth. If a preposition is encountered in the sentence, the parser calls the prepositional phrase (PP) rule.

A NP transition network is shown as the second diagram in Figure 1, where it starts in the initial state (4) and moves to state (5) if it finds a determiner or an adjective, or on to state (6) when a noun is found. The loops for ADJ and PP indicate that more than one adjective or prepositional phrase can occur. Note that the PP rule can in turn call a NP rule, resulting in a nested structure. An example of an analyzed noun phrase is shown in Figures 2 and 3.



GRAMMAR

S → NP VP
NP → (DET) (ADJ*) N (PP*)
PP → PREP NP
VP → VTRAN NP

Figure 1. A Transition Network for a Small Subset of English. Each diagram represents a rule for finding the corresponding word pattern. Each rule can call on other rules to find needed patterns.

After Graham (1979, p214.)

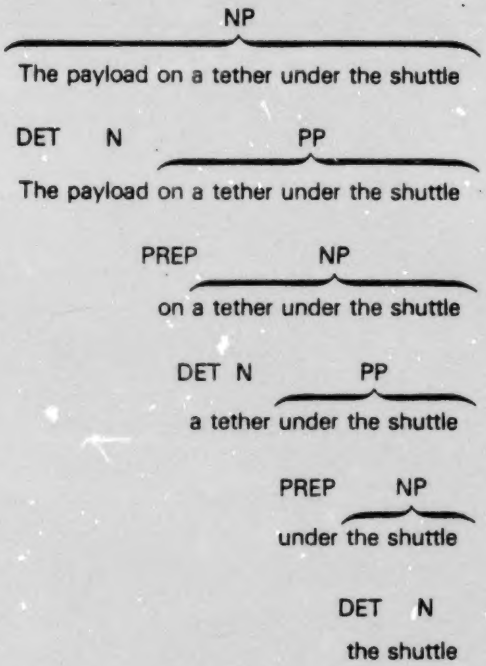


Figure 2. Example Noun Phrase Decomposition.

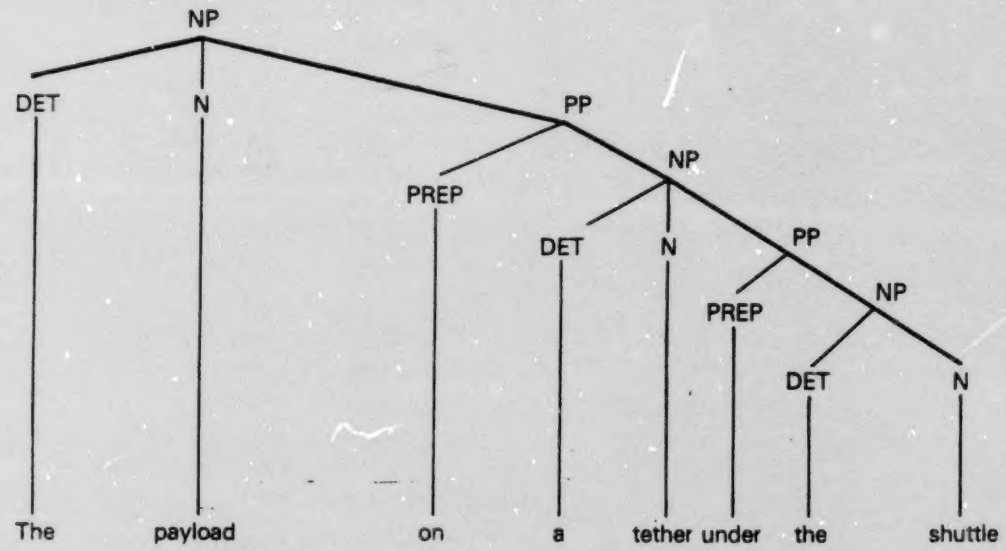


Figure 3. Parse Tree Representation of the Noun Phrase Surface Structure.

As the transition networks analyze a sentence, they can collect information about the word patterns they recognize and fill slots in a frame associated with each pattern. Thus, they can identify noun phrases as singular or plural, whether the nouns refer to persons and if so their gender, etc., needed to produce a deep structure. A simple approach to collecting this information is to attach subroutines to be called for each transition. A transition network with such subroutines attached is called an "augmented transition network," or ATN. With ATN's, word patterns can be recognized. For each word pattern, we can fill slots in a frame. The resulting filled frames provide a basis for further processing.

3. Other Parsers

Other parsing approaches have been devised, but ATN's remain the most popular syntactic parsers. ATN's are top-down parsers in that the parsing is directed by an anticipated sentence structure. An alternative approach is bottom-up parsing, which examines the input words along the string from left to right, building up all possible structures to the left of the current word as the parser advances. A bottom-up parser could thus build many partial sentence structures that are never used, but the diversity could be an advantage in trying to interpret input word strings that are not clearly delineated sentences or contain ungrammatical constructions or unknown words. There have been recent attempts to combine the top-down with the bottom-up approach for NLP in a similar manner as has been done for Computer Vision (see, e.g., Gevarter, 1982).

For a recent overview of parsing approaches see Slocum (1981).

I. Semantics, Parsing and Understanding

The role of syntactic parsing is to construct a parse tree or similar structure of the sentence to indicate the grammatical use of the words and how they are related to each other. The role of semantic processing is to establish the meaning of the sentence. This requires facing up to all the cantankerous ambiguities discussed earlier.

In natural languages (unlike restricted languages, e.g., semantic grammars) it is often difficult to parse the sentences and hook phrases into the proper portion of the parse tree, without some knowledge of the meaning of the sentence. This is especially true when the discourse is ungrammatical. Therefore, it has been suggested that semantics be used to help guide the path of the syntactic parser (see, for example, Charniak, 1981). For that case, syntax presses ahead as far as it can and then hands off its results to the semantic portion to disambiguate the possibilities. Woods (1980) has extended ATN grammars for this purpose. Barr and Feigenbaum (1981, p. 257) indicate that present language understanding systems are indeed tending toward the use of multiple sources of knowledge and are intermixing syntactics and semantics.

Charniak (1981) indicates that there have been two main lines of attack on word sense ambiguity. One is the use of discrimination nets (Reiger and Small, 1979) that utilize the syntactic parse tree (by observing the grammatical role that the word plays, such as taking a direct object, etc.) in helping to decide the word sense. The other approach is based on the frame/script idea (used, e.g., for story comprehension) that provides a context and the expected sense of the word (see, e.g., Schank and Abelson, 1977).

Another approach is "preference semantics" (Wilks, 1975) which is a system of semantic primitives through which the best sense in context is determined. This system uses a lexicon in which the various senses of the words are defined in terms of semantic primitives (grouped into entities, actions, cases, qualifiers, and type indicators). Representation of a sentence is in terms of these primitives which are arranged to relate agents, actions and objects. These have preferential relations to each other. Wilks' approach finds the match that best satisfies these preferences.

Charniak indicates that the semantics at the level of the word sense is not the end of the parsing process, but what is desired is understanding or comprehension (associated with pragmatics). Here the use of frames, scripts and more advanced topics such as plans, goals, and knowledge structure (see, e.g., Schank and Riesbeck, 1981) plays an important role.

J. NLP Systems

As indicated below, various NLP systems have been developed for a variety of functions.

1. Kinds

a. Question Answering Systems

Question answering natural language systems have perhaps been the most popular of the NLP research systems. They have the advantage that they usually utilize a data-base for a limited domain and that most of the user discourse is limited to questions.

b. Natural Language Interfaces (NLI's)

These systems are designed to provide a painless means of communicating questions or instructions to a complex computer program.

c. Computer-Aided Instruction (CAI)

Arden (1980, p. 465) states:

> One type of interaction that calls for ability in natural languages is the interaction needed for effective teaching machines. Advocates of computer-aided instruction have embraced numerous schemes for putting the computer to use directly in the educational process. It has long been recognized that the ultimate effectiveness of teaching machines is linked to the amount of intelligence embodied in the programs. That is, a more intelligent program would be better able to formulate the questions and presentations that are most appropriate at a given point in a teaching dialog, and it would be better equipped to understand a student's response, even to analyze and model the knowledge state of the student, in order to tailor the teaching to his needs. Several researchers have already used the teaching dialogue as the basis for looking at natural languages and reasoning. For example, the SCHOLAR system of Carbonell and Collins tutors students in geography, doing complex reasoning in deciding what to ask and how to respond to a question. Meanwhile, SOPHIE teaches electronic circuits by integrating a natural-language component with a specialized system for simulating circuit behavior. Although these systems are still too costly for general use, they will almost certainly be developed further and become practical in the near future.

d. Discourse

Systems that are designed to understand discourse (extended dialogue) usually employ pragmatics. Pragmatic analysis requires a model of the mutual beliefs and knowledge held by the speaker and listener.

e. Text Understanding

Though Schank (see Schank and Riesbeck, 1981) and others have addressed themselves to this problem, much more remains to be done. Techniques for understanding printed text include scripts and causative approaches.

Arden (1980, pp. 465-466) states:

> To understand a text, a system needs not only a knowledge of the structure of the language but a body of "world knowledge" about the domain discussed in the text. Thus a comprehensive, text-understanding system presupposes an extensive reasoning system, one with a base of common-sense and domain-specific knowledge.
>
> The problem of "understanding" a piece of text does, however, serve as a basic framework for current research in natural languages. Programs are written which accept text input and illustrate their understanding of it by answering questions, giving paraphrases, or simply providing a blow-by-blow account of the reasoning that goes on during the analysis. Generally, the programs operate only on a small preselected set of texts created or chosen by the author for exploring a small set of theoretical problems.

f. Text Generation

There are two major aspects of text generation, one is the determination of the content and textual shape of the message, the second is transforming it into natural language. There are two approaches for accomplishing this. The first is indexing into canned text and combining it as appropriate. The second is generating the text from basic considerations. One need for text generation results from the situation in which information sources need to be combined to form a new message. Unfortunately, simply adjoining sentences from different contexts usually produces confusing or misleading text. Another need for text generation is for explanations of Expert System actions. Text generation will become particularly important as data bases gradually shift to true knowledge bases where complex output has to be presented linguistically. McDonald's thesis (1980) provides one of the most sophisticated approaches to text generation.

g. System Building Tools

Recently, computer languages and programs especially designed to aid in building NLP systems have begun to appear. An example is OWL developed at MIT as a semantic network knowledge representation language for use in constructing natural language question answering systems.

2. Research NLP Systems

Until recently, virtually all of the NLP systems generated were of a research nature. These NLP systems basically were aimed at serving five functions:

a. Interfaces to Computer Programs
b. Data Base Retrieval
c. Text Understanding
d. Text Generation
e. Machine Translation

A few of the more prominent systems are briefly reviewed in this section.

a. Interfaces to Computer Programs

One of the most important early NLP systems, SHRDLU, was a complete system combining syntactic and semantic processing. This system, designed as an interface to a research Blocks World simulation, is described in Table IIa.

SOPHIE (Table IIb), a Computer-Aided Instruction (CAI) system, made use of a semantic grammar to parse the input and to provide instruction based on a simulation of a power supply circuit.

TDUS (Table IIc) uses a procedural network (which encodes basic repair operations) to interpret a dialog with an apprentice engaged in repair of an electro-mechanical pump.

b. Natural Language Interfaces to Large Data Bases

One of the important and prominent research areas for NLP is intelligent front ends to data base retrieval systems. LUNAR (Table IId) is one of the most often cited early systems. It utilized a powerful ATN syntactic parser which passed on its results to a semantic analyzer.

PLANES (Table IIe) was a system designed as a front end to the Navy's database of maintenance and flight records for all naval aircraft. This semantic-grammar-based system ignores the sentence's syntax, searching instead for meaningful semantic constituents by using ATN subnets. These subnets include PLANETYPE, TIME PERIOD, ACTION, etc.

ROBOT (Table IIf) uses an ATN syntactic parser followed by a semantic analyzer to produce a formal query language representation of the input sentence. ROBOT has proved to be very versatile.

LIFER/LADDER (Table IIg) uses patterns or templates to interpret sentences. It employs a semantic (pragmatic) grammar, which greatly simplifies the interpretation. Can handle ellipses and pronouns.

c. Text Understanding

SAM (Table IIh) is a research system that attempts to understand text about everyday events. Knowledge is encoded in frames called scripts. SAM uses an English to Conceptual Dependency parser to produce an internal representation of the story.

PAM (Table IIi) is one offspring of SAM. PAM understands stories by determining the goals that are to be achieved in the story. It then attempts to match actions of the story with methods that it knows will achieve the goals.

d. Text Generation

Winograd (1983) indicates that the difficult problems in generation are those concerned with meaning and context rather than syntax. Thus, until recently, text generation has been mostly an outgrowth of portions of other NLP systems.

e. Machine Translation

Though machine translation was the first attempt at NLP, early failures resulted in little further work being done in this area until recently.

f. Current Research NLP Systems

Table III lists NLP Systems currently being researched.

3. Commercial Systems

The commercial systems available today together with their approximate prices are listed in Table IV. Several of these systems are derivatives of the research NLP systems previously discussed.

TABLE IIa. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
SHRDLU M.I.T. (Winograd, T., 1972) Nat. Lang. Interface to manipulate Blocks World	• Combines syntactic and semantic analysis with a body of world knowledge about a limited domain to provide a NLI to deal with manipulating blocks in a simulation of an artificial "Blocks World." • Starts the analysis of a user's sentence by syntactically parsing a meaningful portion of the sentence. Then semantic routines are called to analyze the unit. The definitions of words in the dictionary are in the form of procedures (procedural semantics) to analyze the unit. These procedures set semantic markers of possible relations to other words. If there are no semantic objections, the syntactic parser continues, otherwise it will try another parse. • Facts are expressed in First Order Predicate Logic. Verifies hypotheses by theorem-proving. • Generates text by "fill in the blank" and stored response patterns. • Heuristically uses pronouns for noun phrases to reduce the stilted nature of the text response. • Type B System	• One of the first systems to deal simultaneously with many sophisticated issues of NLP: —parsing —semantics —references to previous discourse —knowledge representation —problem solving	• Assumes it knows everything about the world. • Assumes world is logical, simple, small and closed. • Required familiarization by user to use it successfully. • Was a prototype that proved to be non-portable and non-extensible and is no longer in use.

TABLE IIb. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
SOPHIE (Sophisticated Instructional Environment) (Brown and Burton, 1975).	• Incorporated a simulation of a power supply circuit to test student suggestions. • Employed a semantic grammar using constitutents like: Request, Fault, Instrument, Node/Name, and Junction/ Type.	• Could run simulations, abstract them and use the results. • Responded in a few seconds. • Could skip words that did not match the grammar rule.	• Skipping words might change meaning of sentence significantly. • The system orgar.ization restricts the system tc only this limited domain.
BBN C.A.I. in Electronic Trouble Shooting.	• The semantic grammar worked much like a syntactic parser, but nodes in resulting parse tree were meaningful semantic units. • Grammar operated top-down in a recursive fashion. • Each grammar rule was a LISP procedure that generated a semantic representation of a subtree in the parse. • Type A + System.	• Very successful and robust.	

TABLE IIc. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
TDUS (Task Oriented Dialogue System) SRI (Robinson, 1980) Interactive Dialog in context. Guide repair operation on electromechanical equipment.	• Goal was to follow the context as an apprentice moved from task to task and respond successfully to his remarks and requests for guidance. • Various tasks to be performed were encoded in procedural networks—an extension of standard network formalisms to allow encoding of quantified information and information about processes. • Uses procedural network to interpret dialog. • Assumes that referential statements refer to objects salient in the current subtask or higher in the task hierarchy. Uses context and discourse to identify objects referred to by definite noun phrases. • Type B+ System.	• Understands contexts, so it can interpret remarks such as "should," "done it," etc. • Can follow particular instantiations of actions. • Realizes the program does not know all things. (Does not operate on "closed world" assumption). • Uses procedural network system to infer unstated intermediate steps.	• Little understanding of the goals and motivations of the apprentice.

TABLE IId. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
LUNAR BBN (Woods, 1973) Natural Language Interface to Moon Rocks Data Base.	• Simplified Data Base —Only a small vocabulary (3500 words) required for moon rock data base. —LUNAR data base encoded in the data base query language. —Seven data domains. Sets of data elements that could be members of each domain were mutually exclusive. • Used a powerful ATN syntactic parser. • Parsed sentence sent on to the semantic program for translation into a query. The resulting query was then executed. • Semantic analyzer gathers information from verbs and their cases, nouns, noun modifiers and determiners to build the data base query. The query is built in terms of the conceptual primitives of the data base. Uses rules to compare the syntactic structure of the question with a syntactic template. If they match, the semantic part of the rule is added to the developing query. • Type B – System.	• Can handle anaphoric references (pronoun references to previous phrases). • Could handle 90% of the questions posed to LUNAR by geologists. • Overall formulation so clean and neat that it has since been used for most parsing and language understanding systems. (Waltz, 1981, p.10).	• As ATN and semantic analyzer are separate, the semantic analyzer must grope thru parsed errors such as prepositional phrases being attached at the wrong point in the parse tree. • Utterances were limited to strict data base inquiries. • Based on a "closed world" viewpoint. • Proved to be non-portable and non-extensible. No longer in use.

TABLE IIe. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
PLANES/JETS (Programmed Language-based Enquiry Sys.) M.I.T. (Waltz, D.L., 1975) Natural Language Interface to a Large Data Base	• Data base is the Navy's 3-M relational data base which holds the maintenance and flight records for all naval aircraft. • Ignores syntax. Assumes that all inputs are in the form of requests that it turns into formal language query expressions. • Uses a semantic grammar. It looks for semantic constituents by doing a left to right scan of the user's sentence. Semantic constituents include items which belong to PLANETYPE, TIMEPERIOD, MALFUNCTION CODE, HOW MANY, ACTION, etc. • Uses an ATN parser. The top level calls various subnets to analyze the input for semantic constituents. • Utilizes concept case frames which are strings of constituents of reasonable queries. • After application of the concept case frames, the resulting semantic constituents are passed along to the query generator. • Type A System.	• Can handle ellipses and pronouns. • Can deal with some nongrammatical sentences. • Asks for a rephrase if it doesn't understand.	• Relatively inefficient, could benefit from a look ahead. A look ahead could result in an order of magnitude reduction in number of arcs tested in the parse of a sentence. • Problems with word sense selection and modifier attachment. PLANES relies too heavily on its particular world of discourse for eliminating problems of word sense selection. • In a 1980 test, PLANES understood about 2/3 of queries correctly. Could be made into a useful practical program with further work.

TABLE IIf. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
ROBOT/INTELLECT Dartmouth (Harris, 1977) Data Base Question Answering System.	• Uses an ATN syntactic parser (with backtracking) followed by semantic analysis to produce a formal query language representation of the input sentence. • Handles a large vocabulary by building an inverted file of data element names indicating the data domains in which each name occurs. In addition, the inverted file contains words and phrases that are interpreted as data element names. • A dictionary of common English words is also included. • If two meanings of the inquiry appear likely, and only one returns hits, that one is interpreted to be the appropriate one. • Type A System.	• INTELLECT is one of the first N.L. Data Base Query systems to be available commercially. • Can handle idioms via special mechanisms. • Can adapt INTELLECT to a new data base in approximately one week. • Can handle some pronouns and ellipses.	• Does not consider context except to disambiguate pronouns and ellipses.

TABLE IIg. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
LADDER (Language Access to Distributed Data with Error Recovery). SRI (Hendrix et al., 1978). Natural Language Data Base Query.	• Application of LIFER parser. • Uses patterns or templates to interpret sentences. Associates a function with each pattern. • Uses a Semantic (pragmatic) grammar and associated functions to implicitly encode knowledge about language and the world. The grammar contains much information about the particular data base being queried. • Type A System.	• Can correct spelling. • Can handle ellipsis. • Can interpret pronouns. • Can deal with large and complex data bases, e.g., in Naval Ship DB has dealt with: —100 fields in 14 files —records of 40,000 ships. • Can answer certain questions based upon its own N.L. processing system. • Can be taught synonyms. • Can be taught new syntactic constructions. • Can accept a defined input sentence as equivalent to a whole set of questions.	• Conversation is limited strictly to questions about a small domain. • Can't deal with logically complex notions: —disjunction —quantification —implication —causality —possibility • Closed-world viewpoint Acts as if it was dealing with a world —containing a fixed number of objects and relationships between them —with objects and relationships being immutable.

TABLE IIh. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
SAM (Script Analyzer Mechanism) Yale (Shank et al., 1975). Understands events using prototype descriptions of them.	• Knowledge of prototypical events is encoded in frames called scripts. • Utilizes a domain dictionary. The first word sense that satisfies the local context (as provided by the script) is selected. (Thus scripts are a convenient means for interpreting words with multiple senses). • Understands stories by fitting them to a script in a three part process: 1. Parser generates a conceptual dependency (CD) representation for each sentence. 2. A script applier (APPLY) gives it a set of verb-senses to use once a script is identified. Then it checks to see if the CD sentence representation matches the current script or any other script in the data base. If this matching is successful, APPLY makes a set of predictions about likely inputs to follow. Any steps in the current script that were left out in the story, are filled in. 3. A memory module takes resultant references to people, places, things, etc. and fills in information about them. • Type B System.	• Can produce a summary of the story (in several different languages) or answer questions about it. • Can produce paraphrases of the story and make intelligent inferences from it. • Can infer missing information by using the script.	• Knowledge is primarily about everyday world, rather than about natural language. • Only a single object can serve the role of a player or a prop. • Scripts follow a linear sequence—can't deal with alternative possibilities. • Difficult to determine which scripts are appropriate for a given story.

TABLE IIi. Natural Language Understanding Systems.

System/Use	Approach	Capabilities	Limitations
PAM Yale (Wilensky, 1978) Story Understanding	• Understands stories by determining the goals that are to be achieved in the story. PAM then attempts to match actions of the story with methods that it knows will achieve goals. • Has a knowledge base of plans and themes. • A plan is a set of actions and subgoals for accomplishing the main goal. • Themes are basic situations encountered in life, such as "love." • Program starts by converting written text into CD representation (as in SAM). • Goals of an actor are determined in the following ways. —noting them explicitly in story. —using plans, establishing them as subgoals to a known goal. —inferring them from a theme noted in the story. • Type B-C System.	• Can summarize a story. • Can answer questions about goals and actions of the characters. • Can extend SAM to stereotyped situations.	• A great deal of inference can be required by PAM to establish the goals and subgoals of the story from the input text. • Much must be known about the nature of the story to be sure that the needed stored plans and themes are available.

TABLE III. Current Research NLP Systems.

System	Purpose	Developer	Comments
EUFID (End-User Friendly Interface to Data)	NLI to DBMS	System Development Corp. Santa Monica, California	• Application Independent. • Uses an Intermediate Language as the output of the NL analysis system. Then translates from this to the target DBMS query language.
ASK (A Simple Knowledgeable System)	NLI for users creating own data base	CA Inst. of Technology Pasadena, California	• Uses a limited dialect of English. • Develops a Semantic Net with nodes limited to Classes, Objects, Attributes and Relations, and the appropriate corresponding arcs.
NLP + DBAP	NLI to a DB	Bell Labs Murray Hill, New Jersey	• Consists of two parts, a Natural Language Processor (NLP) and a Data Base Application Program (DBAP). • The NLP is general purpose language processor which builds a formal representation of the input. The DBAP is an algorithm which builds a query in an augmented relational algebra from the output of the NLP. • System is portable and said to be very robust.

TABLE III. Current Research NLP Systems. (continued)

System	Purpose	Developer	Comments
IR-NLI (Internal Representation -NLI)	NLI for an on-line information retrieval system.	U. of Udine Udine, Italy	• Utilizes a base of expert knowledge, which concerns the evaluation of the user's requests, the management of the research interview, the selection of search strategy and the scheduling of the lower level modules: UNDERSTANDING and DIALOGUE, REASONING and FORMALIZER. • The UNDERSTANDING and DIALOGUE Module translates the user's requests into a basic formal internal representation.
TEAM: (Transportable English Access Data Manager)	Transportable NLI	SRI Inter. Menlo Park, California	• Has three major components: • An acquisition component • The DIALOGIC Language System • Data-Access Component. • Utilizes the acquisition component to obtain (via an interactive dialogue with the DB management personnel) the information required to adapt the system to a particular DB. • Translates English query into a DB query in two steps — The DIALOGIC system constructs a logical representation of the query. — The data-access component translates the logic form into a formal DB query.

TABLE III. Current Research NLP Systems. (continued)

System	Purpose	Developer	Comments
NOMAD	Text Understanding	AI Project U. of California Irvine, California	• Uses internal syntactic and semantic expectations to understand unedited naval ship-to-shore messages. • Utilizes a large data base of domain specific knowledge. • Outputs a corrected well-formed English translation of the message. • Utilizes knowledge of syntax, semantics, and pragmatics at all stages of the understanding process to cope with errors.
(Automated Analysis of Descriptive Texts)	Text Understanding	U. of Strathclyde Glasgow, Scotland	• Instantiates domain-dependent hierarchical frame-like structures (written in PROLOG) by identifying key words and using a domain dictionary.
BEDE	Machine Translation	U. of Manchester England	• Analyzes source text and translates it into an intermediate (Interlingua) language. Then synthesizes target language text from this. • Allows only a controlled vocabulary and a restricted syntax, with the aim of microprocessor-based MT.
(English-Japanese MT)	Machine Translation	Kyoto U. Japan	• Uses Montague Grammar to generate an intermediate representation of meaningful semantic relations in a functional logical form. Converts the logical form to a conceptual phrase structure form associated with Japanese.

TABLE III. Current Research NLP Systems. (continued)

System	Purpose	Developer	Comments
LRC MT	Machine Translation	U. of Texas for Siemens Munich, W. Germany	• Employs a phrase-structure (PS) grammar augmented by lexical controls. • Utilizes over 400 PS rules describing the source language (German) and nearly 10,000 lexical entries in each of two languages (German and the target language—English). • Uses an all-paths, bottom-up parser. • Uses special procedures to cope with ungrammatical input.
(Not Named NLP System)	NLI to an inferencing KB	Hewlett Packard Palo Alto, California	• Systems main components are: — A Generalized Phrase Structure Grammar — A top-down parser — A logic transducer that outputs a first-order logical representation. — A "disambiguator" that uses so tal information to convert logical expressions into the query language for HIRE (a relational data base).
KLAUS (Knowledge-Learning and -Using System)	Computer acquisition of a model of a domain of interest by being instructed in English.	SRI International Menlo Park, California	• Uses SRI's DIALOGIC NLP System to translate English sentences into logical representations of their literal meaning in the context of the utterance. • KLAUS is a DARPA-sponsored long-term research project to develop techniques for facilitating the acquisition of knowledge by computer.

TABLE III. Current Research NLP Systems. (continued)

System	Purpose	Developer	Comments
TEXT	Text Generation	U. of Pennsylvania Phila., Pennsylvania	• Schemas which encode aspects of discourse structure, are used to guide the discourse process. • A focusing mechanism monitors the use of the schemas, providing constraints on what can be said at any point. • On the basis of the input question, semantic processes produce a relevant knowledge pool. A partially ordered set of rhetorical techniques are selected as appropriate for the pool. A message is generated by matching propositions in the pool to the associated rhetorical techniques.
EPISTLE	Text Understanding and Text Generation	IBM C.S. Dept. Yorktown Hts., New York	• Utilizes an augmented phrase structure grammar. • The core grammar consists at present of a set of 300 syntax rules. • Ambiguity is resolved by using a metric that ranks alternative parses. • A "fitted-parse" technique is used to produce reasonable approximate parses to ungrammatical inputs. • Uses an on-line dictionary with about 130,000 entries.

TABLE III. Current Research NLP Systems. (concluded)

System	Purpose	Developer	Comments
TOPIC	Automatic Text Condensation	U. of Constance Infor. Sci. Dept. West Germany	• Uses frame-oriented knowledge representation models. • Utilizes "interacting word experts" approach to aid in textual parsing.
KAMP	NL Generation	SRI International Menlo Park, California	• Plans NL utterances, starting with a high-level description of the speaker's goals. • The heuristic plan generation process is by a NOAH-like hierarchical planner, and verified by a first order logic theorem prover. • The planner uses knowledge about the different subgoals to be achieved and linguistic rules about English to produce utterances that satisfy multiple goals.

TABLE IV. Some Commercial Natural Language Systems.

System	Organization	Purpose	Comments
INTELLECT (Derivative of ROBOT) $50K/system (also distrib- uted as ON- LINE ENGLISH and GRS Executive)	Artificial Intelligence Corp. Waltham, Massachusetts (Culliane) (Information Sciences)	NLI for Data Base Retrieval. (Other extensions underway).	• Several hundred systems sold. • Takes about 2 weeks to implement for a new data base. • Written in PL-1. • Available for mainframes.
PEARL (Based on SAM and PAM) $250K/ system	Cognitive Systems New Haven, Connecticut	Custom NLI's. The first system— Explorer—is an interface to an existing map gen- erating system. Others are interfaces to data bases.	• Large start-up cost in building the knowledge base. • Several systems have been, and are being, built. • Written in LISP.
STRAIGHT TALK (Derivative of LIFER) $660/system	Dictaphone, Written by Symantec Sunnyvale, California	Highly portable NLI for DBMS for micro- computers.	• Written in PASCAL. Designed to be very compact and efficient. Available about Nov. 1983. • User customized.
SAVVY $250/system	SAVVY Marketing International Sunnyvale, California	System Interface for micro-computers.	• Not linguistic. User adaptive (best fit) pattern matching to strings of characters. • Released 3/82. • User customized.

TABLE IV. Some Commercial Natural Language Systems. (continued)

System	Organization	Purpose	Comments
Weidner System $16K/language direction	Weidner Communications Corp. Provo, Utah	Semi-Automatic Natural Language Translation.	• Linguistic approach. Written in FORTRAN IV. • Translation with human editing is approximately 1600 words/hr (up to eight times as fast as human alone). • Approx. 20 sold by end of 1982, mainly to large multi-national corporations.
ALPS	ALPS Provo, Utah	Interactive Natural Language Translation.	• Linguistic Approach. • Uses a dictionary that provides the various translations for technical words as a display to human translator, who then selects among the displayed words.
NLMENU	Texas Instruments, Inc. Dallas, Texas	NLI to Relational Data Bases.	• Menu Driven NL Query System. • All queries constructed from menu fall within linguistic and conceptual coverage of the system. Therefore, all queries entered are successful. • Grammars used are semantic grammars written in a context-free grammar formalism. • Producing an interface to any arbitrary set of relations is automated and only requires a 15-30 minute interaction with someone knowledgeable about the relations in question. • System will be available late in 1983 as a software package for a microcomputer.

K. State of the Art

It is now feasible to use computers to deal with natural language input in highly restricted contexts. However, interacting with people in a facile manner is still far off, requiring understanding of where people are coming from—their knowledge, goals and moods.

In today's computing environment, the only systems that perform robustly and efficiently are Type A systems—those that do not use explicit world models, but depend on key word or pattern matching and/or semantic grammars. In actual working systems, both understanding and text generation, ATN-like grammars can be considered the state of the art.

L. Problems and Issues

1. How People Use Language

Many of the issues in natural language understanding center around the way people use language. Given speech acts can serve many purposes, depending on the goals, intentions and strategies of the speaker. Thus, methods for determining the underlying motivation of a speech act is a major issue. Another issue is understanding how humans process language—both in forming output and in interpreting input.

It also appears that knowledge-based inference is essential to natural language understanding, as language just provides abbreviated cues that must be fleshed out using models and expectations resident in the receiver. Finally, we do not even have a good handle on what it means to understand language and what is the relation between language and perception.

2. Linguistics

A major issue in NLP is how to resolve ambiguities in word meanings to determine their appropriate sense in the current context. A complementary problem is dealing with novel language such as metaphors, idioms, similes and analogies.

Syntactic ambiguity is a common source of trouble in natural language processing. Where to attach modifying clauses is one problem. However even handling adverbial modifiers has proved difficult.

Another major issue is pragmatics—the study of language in context. Arden (1980, p. 474) notes:

> Many of the issues discussed under frame systems are pertinent to pragmatic issues. The prototypes stored in a frame system can include both the prototypes for the domain being discussed and those related to the conversational situation. In a travel-planning system, then, a user responds to the question, "What time do you want to leave?" with the answer: "I have to be at a meeting by 11." In planning an appropriate flight, the system makes assumptions about the relevance of the answer to the question.
>
> This aspect of language is one that is just beginning to be dealt with in current systems. Although most large systems in the past had specialized ways of dealing with a subset of pragmatic problems, there is as yet no theoretical approach. As people look to interactive systems for teaching and explanation, however, it seems likely that this will be the major focus of research in the 1980's.

3. Conversation

In the area of everyday conversation, the real world is extensive, complex, largely unknown and unknowable. This is quite different from the closed world of many of the research NLP systems.

"A major problem for NLP systems is following the dialogue context and being able to ascertain the references of noun phrases by taking context into account." (Hendrix and Sacerdoti, 1981, p. 330)

Another major problem is understanding the motivation of the participants in the discourse in order to penetrate their remarks. As conversational natural-language communication between individuals is dependent on what the participants know about each other's knowledge, beliefs, plans, and goals, methods for developing and incorporating this knowledge into a computer are major issues.

4. Processor Design

"While many specific problems are linguistic, . . . many important problems are actually general AI problems of representation and process organization." (Arden, 1980, p. 409)

A major issue in the design of a NLP system is choosing the tradeoffs between capability, efficiency and simplicity. Also at issue are the language constructs to be handled, generality, processing time and costs. The choice of the overall architecture of the system and the grammar to be used is a major design decision for which there are as yet no general criteria.

Though all natural-language processing systems contain some sort of parser, the practical design of applications of grammar to NLP has proved difficult. The design of the parser in both theory and implementation is a complex problem. Also at issue is the top-down (ATN-like) approach to parsing versus bottom-up and combined approaches. In addition, how best to utilize knowledge sources (phonemic, lexical, syntactic, semantic, etc.) in designing a parser and a system architecture remains a major issue.

A problem with the ATN parser approach, with its heavy dependence on syntax, is how can it be adapted to handle ungrammatical inputs. Though considerable progress has been made, there is as yet no clear solution. INTELLECT (a commercial ATN-based system) handles ungrammatical constructions by relaxing syntactic constraints. IBM's Epistle System (Jensen and Heidorn, 1983) uses a fitting procedure to ungrammatical inputs to produce a reasonable approximate parse. Semantic grammars and expectation-driven systems have an advantage in overcoming ungrammatical inputs.

Another major issue is: Is it appropriate to keep the semantic analysis separate from the syntactic analysis, or should the two work interactively? (see Charniak, 1981)

Also, is it necessary in NL translating or understanding to utilize an intermediate representation, or can the final interpretation be gotten at more directly? If an intermediate representation is to be used, which one is best? What is the appropriate role of primitive concepts (such as found in case systems or conceptual dependency) in natural language processing?

How can we make restricted natural language more palatable to humans? A major problem is the negative expectations created in the mind of a naive user, when a system doesn't understand an input sentence. Naive users have difficulty distinguishing between the limitations in a system's conceptual coverage and the system's linguistic coverage. A related problem is the system returning a null answer. This may mislead the user as an answer may be null for many reasons. Another problem is insuring a sufficiently rapid response to user inputs.

One common problem with real systems is stonewalling behavior—the system not responding to what the user is really after (the user's goal) because the user hasn't suitably worded the input.

Some of the important problems and issues have to do with knowledge representation:

—Which knowledge representation is appropriate for a given problem?

—How to represent such things as space, time, events, human behavior, emotions, physical mechanisms and many processes associated with novel language?

—How can common sense and plausibility judgement (is that meaning possible?) be represented?

—How should items in memory be indexed and accessed?

—How should context be represented?

—How should memory be updated?

—How to deal with inconsistencies?

—How can we make the representations more precise?

—How can we make the system learn from experience so as to build up the necessary large knowledge needed to deal with the real world?

—How can we build useful internal representations that correspond to 3D models, from information provided by natural language?

NLP usually takes the sentence as the basic unit to be analyzed. Assigning purpose and meaning to larger units has proved difficult. The NRL Conceptual Linguistics Workshop (1981) concluded that "Concept extraction was the most difficult task examined at the workshop. Success depends on the adequacy of the situation-context representation and the development of more sophisticated models of language use."

NLP has always pushed the limits of computer capability. Thus a current problem is designing special computer architectures and processors for NLP.

5. Data Base Interfaces

Hendrix and Sacerdoti (1981, pp 318, 350) point out two problems particularly associated with data base interfaces:

> (1). The need to understand context throws considerable doubt on the idea of building natural-language interfaces to systems with knowledge bases independent of the language processing system itself.
>
> (2). One of the practical problems currently limiting the use of NLP systems for accessing data bases is the lack of trained people and good support tools for creating the knowledge structures needed for each new data base.

6. Text Understanding

Text understanding systems have encountered problems in achieving practicality, both in terms of extending the knowledge of the language and in providing a sufficiently broad base of world knowledge. The NRL Conceptual Linguistics Workshop (1981) concluded that "Current systems for extracting information from military messages use the key word and key phrase methods which are incapable of providing adequate semantic representation. In the immediate future, more general methods for concept extraction probably will work well only in well defined subfields that are carefully selected and painstakingly modeled."

SRI and the National Library of Medicine have text understanding systems in the research stage. SRI handcodes logic formulas that describe the content of a paragraph. Queries are matched against these paragraph descriptions.

M. Research Required

Current research in natural language processing systems includes machine translation, information retrieval and interactive interfaces to computer systems. Important supporting research topics are language and text analysis, user modeling, domain modeling, task modeling, discourse modeling, reasoning and knowledge representation.

Much of the research required (as well as the research now underway) is centered around addressing the problems and issues discussed in the following areas:

1. How People Use Language

The psychological mechanisms underlying human language production is a fertile field for investigation. Efforts are needed to build explicit computational models to help explain why human languages are the way they are and the role they play in human perception.

2. Linguistics

Further research is needed on methods for resolving ambiguities in language and for the utilization of context in language understanding.

3. Conversation

Additional work is needed on ways to represent the huge amount of knowledge needed for Natural Language Understanding (NLU).

A great deal of research is needed to give NLU systems the ability to understand not only what is actually said, but the underlying intention as well.

Research is now underway by many groups on explicitly modeling goals, intentions and planning abilities of people. Investigation of script and frame-based systems is currently the most active NLP AI research area.

4. Processor Design

Architectures, grammars, parsing techniques and internal representations needed for NLP systems remain important research areas.

One particularly fertile area is how to best utilize semantics to guide the path of the syntactic parser. Charniak (1981, p 1085) indicates that a relatively unexplored area requiring research is the interaction between the processes of language comprehension and the form of semantic representation used.

Further work is needed on bringing multiple knowledge sources (KS's: syntactic, semantic, pragmatic and contextual) to bear on understanding a natural language utterance, but still keeping the KS's separate for easy updating and modification. Also needed is further work in AI

problem-solving to cope with the problem of finding an appropriate structure in the huge space of possible meanings of a natural language input.

Improved NLU techniques are needed to handle complex notions such as disjunction, quantification, implication, causality and possibility. Also needed are better methods for handling "open worlds," where all things needed to understand the world are not in the system's knowledge base.

Further research is also necessary to aid with a common source of trouble in NLP, that is, dealing with syntactic and semantic ambiguities and how to handle metaphors and idioms.

Finally, the problems of efficiency, speed, portability, etc., discussed in the previous chapter, all are in need of better solutions.

5. Data Base Interfaces

A current research topic is how can data base schemas best be enriched to support a natural language interface, and what would be the best logical structure for a particular data base.

Research is also needed on more efficient methods for compiling a vocabulary for a particular application.

6. Text Understanding

Seeking general methods of concept extraction remains as one of the major research areas in text understanding.

N. Principal U.S. Participants in NLP

*1. Research and Development**

Non-Profit

SRI
MITRE

Universities

Yale U. — Dept of Computer Science
U. of CA, Berkeley — Computer Science Div., Dept of EECS.
Carnegie-Mellon U. — Dept of Computer Science.
U. of Illinois, Urbana — Coordinated Science Lab.
Brown U. — Dept of Computer Science
Stanford U. — Computer Science Dept.
U. of Rochester — Computer Science Dept.
U. of Mass, Amherst — Department of Computer and Information Science
SUNY, Stoneybrook — Dept of Computer Science
U. of CA, Irvine — Computer Science Dept.

*A review of current research in NLP is given in Kaplan (1982).

U of PA — Dept of Computer and Infor. Science
GA Institute of Technology — School of Infor. and Computer Science
USC — Infor. Science Institute.
MIT — AI Lab.
NYU — Computer Science Dept. and Linguistic String Project
U. of Texas at Austin — Dept of Computer Science
Cal. Inst. of Tech.
Brigham Young U. — Linguistics Dept.
Duke U. — Dept of Computer Science
N Carolina State — Dept. of Computer Science
Oregon State U. — Dept of Computer Science

Industrial

BBN
TRW Defense Systems
IBM, Yorktown Heights, N.Y.
Burroughs
Sperry Univac
Systems Development Corp, Santa Monica
Hewlett Packard
Martin Marietta, Denver
Texas Instruments, Dallas
Xerox PARC
Bell Labs
Institute for Scientific Information, Phila., PA
GM Research Labs, Warren, MI
Honeywell

2. *Principal U.S. Government Agencies Funding NLP Research*
ONR (Office of Naval Research)
NSF (Nationai Science Foundation)
DARPA (Defense Advanced Research Projects Agency)

3. *Commercial NLP Systems*
Artificial Intelligence Corp., Waltham, Mass.
Cognitive Systems Inc., New Haven, Conn.
Symantec, Sunnyvale, CA.
Texas Instruments, Dallas, TX.
Weidner Communications, Inc., Provo, Utah
SAVVY Marketing Inter., San Mateo, CA.
ALPS, Provo, UT.

4. Non-U.S.

U. of Manchester, England
Kyoto U., Japan
Siemens Corp. Germany
U of Strathclyde, Scotland
Centre National de la Recherche Scientifique, Paris
U. di Udine, Italy
U. of Cambridge, England
Philips Res. Labs, The Netherlands

O. Forecast

Commercial natural language interfaces (NLI's) to computer programs and data base management systems are now becoming available. The imminent advent of NLI's for micro-computers is the precursor for eventually making it possible for virtually anyone to have direct access to powerful computational systems.

As the cost of computing has continued to fall, but the cost of programming hasn't, it has already become cheaper in some applications to create NLI systems (that utilize subsets of English) than to train people in formal programming languages.

Computational linguists and workers in related fields are devoting considerable attention to the problems of NLP systems that understand the goals and beliefs of the individual communicators. Though progress has been made, and feasibility has been demonstrated, more than a decade will be required before useful systems with these capabilities will become available.

One of the problems in implementing new installations of NLP systems is gathering information about the applicable vocabulary and the logical structure of the associated data bases. Work is now underway to develop tools to help automate this task. Such tools should be available within 5 years.

For text understanding, experimental programs have been developed that "skim" stylized text such as short disaster stories in newspapers (DeJong, 1982). Despite the practical problems of sufficient world knowledge and the extension of language knowledge required, practical tools emerging from these efforts should be available to provide assistance to humans doing text understanding within this decade.

The NRL Computational Linguistic Workshop (1981) concluded that text generation techniques are maturing rapidly and new application possibilities will appear within the next five years.

The NRL workshop also indicated that:

> Machine aids for human translators appear to have a brighter prospect for immediate application than fully automatic translation; however, the Canadian French-English weather bulletin project is a fully automatic system in which only 20% of the translated sentences require minor rewording before public release. An ambitious common market project involving machine translation among six European langauges is scheduled to begin shortly. Sixty people will be involved in that undertaking which will be one of the largest projects undertaken in computational linguistics.* The panel was divided in its forecast on the five year perspective of machine translation but the majority were very optimistic.

*EUROTA—A machine translation project sponsored by the European Common Market—8 countries, over 15 universities, $24 M over several years.

Nippon Telegram and Telephone Corp in Tokyo has a machine translation AI project underway. An experimental system for translating from Japanese to English and vice versa is now being demonstrated. In addition, the recently initiated Japanese Fifth Generation Computer effort has computer-based natural language understanding as one of its major goals.

In summary, natural language interfaces using a limited subset of English are now becoming available. Hundreds of specialized systems are already in operation. Major efforts in text understanding and machine translation are underway, and useful (though limited) systems will be available within the next five years. Systems that are heavily knowledge-based and handle more complete sets of English should be available within this decade. However, systems that can handle unrestricted natural discourse and understand the motivation of the communicators remain a distant goal, probably requiring more than a decade before useful systems appear.

As natural language interfaces coupled to intelligent computer programs become widespread, major changes in our society are likely to result. There is a trend now to replace relatively unskilled white collar and factory work with trained computer personnel operating computer-based systems. However, with the advent of friendly interfaces (and eventually even speech understanding systems and automatic text generation from speech) relatively unskilled personnel will be able to control complex machines, operations, and computer programs. As this occurs, even relatively skilled factory and white collar work may be taken over by these lesser skilled personnel with their computer aids—the experts and computer personnel moving on to develop new programs and applications.

The outcome of such a revolution cannot be fully predicted at this time, other than to suggest that much of the power of the computer age will become available to everyone, requiring a rethinking of our national goals and life styles.

P. Further Sources of Information

1. Journals

- *American Journal of Computational Linguistics*—published by the major society in NLP, the Association for Computational Linguistics (ACL).
- *SIGART Newsletter*—ACM (Association for Computing Machinery).
- *Artificial Intelligence*
- *Cognitive Science*—Cognitive Science Society
- *AI Magazine*—American Association for AI (AAAI)
- *Pattern Analysis and Machine Intelligence*—IEEE
- *International Journal of Man Machine Interactions*

2. Conferences

- Computational Linguistics (COLING)—held biannually. Next one is in July 1984 at Stanford Universi y.
- International Joint Conference on AI (IJCAI)—biannual. Current one in Germany, August 1983.
- ACL Annual Conference.

- AAAI annual conferences.
- ACM conferences.
- IEEE Systems, Man & Cybernetics Annual Conferences.
- Conference on Applied Natural Language Processing. Sponsored jointly by ACL & NRL—Feb. 1983 in Santa Monica, CA.

3. Recent Books

- Winograd, T., *Language as a Cognitive Process, Vol I, Syntax,* Reading, Mass: Addison Wesley, 1983.
- Lehnert, W.G. and Ringle, M.H. (eds.), *Strategies for Natural Language Processing,* Hillsdale, N.J. Lawrence Erlbaum, 1982.
- Sager, N., *Natural Language Information Processing,* Reading, Mass: Addison-Wesley, 1981
- Tennant, H., *Natural Language Processing,* New York: Petrocelli, 1981.
- Brady, M., *Computational Approaches to Discourse,* Cambridge, Mass: MIT Press, 1982.
- Joshi, A.K., Weber, B.L. and Sag, I.A. (eds), *Elements of Discourse Understanding,* Cambridge: Cambridge University Press, 1981.
- L. Bolc (ed.), *Natural Language Communication with Computers,* Berlin: Springer-Verlag, 1981.
- L. Bolc (ed.), *Data Base Question Answering Systems,* Berlin: Springer-Verlag, 1982.
- Schank, R.C. and Riesbeck, C.K., *Inside Computer Understanding.* Hillsdale, N.J.: Lawrence Erlbaum, 1981.

4. Overviews and Surveys

- Barr, A and Feigenbaum, E.A., Chapter IV, "Understanding Natural Language," *The Handbook of Artificial Intelligence, Vol I,* Los Altos, CA: W. Kaufmann, 1981, pp 223-322.
- S.J. Kaplan, "Special Section—Natural Language," *SIGART Newsletter,* No. 79, Jan. 1982, pp 27-109.
- Charniak, E., "Six Topics in Search of A Parser: An Overview of AI Language Research," *IJCAI-81,* pp 1079-1087.
- Waltz, D.L., "The State of the Art in Natural Language Understanding," In *Strategies for Natural Language Processing,* W.G. Lehnert and M.H. Ringle (eds), Hillsdale, N.J.: Lawrence Erlbaum, 1982, pp. 3-32.
- Slocum, J., "A Practical Comparison of Parsing Strategies for Machine Translation and Other Natural Language Processing Purposes," Tech. Report NL-41, Dept of C.S., U. of Texas, Aug 1981.
- Hendrix, G. G. and Sacerdoti, E.D., "Natural-Language Processing: The Field in Perspective," *Byte,* Sept. 1981, pp 304-352.

REFERENCES

- Arden, B.W. (ed), *What Can Be Automated? (COSERS),* Cambridge, Mass: MIT Press, 1980.
- Barr, A. and Feigenbaum, E.A., Chapter 4, "Understanding Natural Language," *The Handbook of Artificial Intelligence,* Los Altos, CA: W. Kaufman, 1981, pp 223-321.
- Barrow, H.G., "Artificial Intelligence: State of the Art," Technical Note 198, Menlo Park, CA: SRI International, Oct. 1979.
- Brown, J.S. and Burton, R.R., "Multiple Representations of Knowledge for Tutorial Reasoning." *In Representation of Learning,* D.G. Bobrow and A. Collins (Eds.), New York: Academic Press, 1975.
- Burton, R.R., "Semantic Grammar: An Engineering Technique for Constructing Natural Language Understanding Systems," BBN Report 3453, BBN, Cambridge, Dec. 1976.
- Charniak, E., "Six Topics in Search of a Parser: An Overview of AI Language Research," IJCAI-81, pp 1079-1087
- Charniak, E. and Wilks, Y., *Computational Semantics,* Amsterdam: North Holland, 1976.
- Chomsky, N., *Syntactic Structures,* The Hague: Mouton, 1957.
- DeJong, G., "An Overview of the FRUMP System." In *Strategies for Natural Language Processing,* W.G. Lehnert and M.H. Ringle (eds), Hillsdale, N.J.: Lawrence Erlbaum, 1982, pp 149-176.
- Fillmore, C., "Some Problems for Case Grammar" In R.J. O'Brien (Ed.), *Report of the Twenty-Second Annual Round Table Meeting on Linguistics and Language Studies,"* Wash., D.C.: Georgetown U. Press, 1971, pp. 35-56.
- Finin, T. W., "The Semantic Interpretation of Compound Nominals," Ph.D. Thesis, U. of IL, Urbana, 1980.
- Gawron, J.M. et al., "Processing English with a Generalized Phrase Structure Grammar," *Proc. of the 20th Meeting of ACL,* U. of Toronto, Canada, 16-18 June 1982, pp 74-81.
- Gazdar, G., "Unbounded Dependencies and Coordinate Structure," *Linguistic Inquiry,* 12, 1981, pp. 155-184.
- Gevarter, W.B., *An Overview of Computer Vision,* NBSIR 82-2582, National Bureau of Standards, Wash., D.C., September 1982.
- Gevarter, W.B., *An Overview of Artificial Intelligence and Robotics, Vol. 1,* NBS (in press), 1983.
- Graham, N., *Artificial Intelligence,* Blue Ridge Summit, PA: TAB Books, 1979.
- Hendrix, G. G. and Sacerdoti, E.D., "Natural-Language Processing: The Field in Perspective," *Byte,* Sept. 1981, pp. 304-352.
- Hendrix, G.G., Sacerdoti, E.D., Sagalowicz, D., and Slocum, J., "Developing a Natural Language Interface to Complex Data," ACM *Transactions on Database Systems,* Vol. 3, No. 2, June 1978.

Preceding page blank

• Jensen, K. and Heidorn, G.E., "The Fitted Parse: 100% Parsing Capability in a Syntactic Grammar of English," *Conf. on Applied NLP,* Santa Monica, CA, Feb. 1983, pp. 93-98.

• Kaplan, S.J., (Ed.), "Special Section—Natural Language," *SIGART NEWSLETTER* #79, Jan. 1982, pp. 27-109.

• McDonald, D.B., "Understanding Noun Compounds," Ph.D. Thesis, Carnegie-Mellon U., Pittsburgh, 1982.

• McDonald, D.B., "Natural Language Production as a Process of Decision-Making Under Constraints," Ph.D. Thesis, M.I.T., Cambridge, 1980.

• Nishida, T. and Doshita, S., "An Application of Montague Grammar to English-Japanese Machine Translation," *Proc. of Conf. on Applied NLP,* Santa Monica, Feb. 1983.

• Reiger, C. and Small, S., "Word Expert Parsing," *Proc. of the Sixth International Joint Conference on Artificial Intelligence,* 1979, pp. 723-728.

• Robinson, A.E., et al., "Interpreting Natural Language Utterances in Dialog about Tasks," AI Center TN 210, SRI Inter., Menlo Park, CA, 1980.

• Schank, R.C. and Abelson, R.P., *Scripts, Plans, Goals and Understanding,* Hillsdale, N.J.: Lawrence Erlbaum, 1977.

• Schank, R.C. and Riesbeck, C.K., *Inside Computer Understanding,* Hillsdale, N.J.: Lawrence Erlbaum, 1981.

• Schank, R.C. and Yale AI Project, "SAM—A Story Understander," Research Rept 43, Dept of Comp Sci, Yale U., 1975.

• Slocum, J., "A Practical Comparison of Parsing Strategies for Machine Translation and Other Natural Language Purposes," Ph.D. Thesis, U. of Texas, Austin, 1981.

• Tennant, H., *Natural Language Processing,* New York: Petrocelli, 1981.

• Waltz, D.L., "Natural Language Access to a Large Data Base," In *Advance Papers of the International Joint Conference on Artificial Intelligence,* Cambridge, Mass, MIT, 1975.

• Waltz, D.L., "The State of the Art in Natural Language Understanding," In *Strategies for Natural Language Processing.,* W.G. Lehnert and M.H. Ringle (eds), Hillsdale, N.J.: Lawrence Erlbaum, 1982, pp. 3-32.

• Webber, B.L. and Finin, T.W., "Tutorial on Natural Language Interfaces, Part 1—Basic Theory and Practice," AAAI-82 Conference, Pittsburgh, PA, Aug. 17, 1982.

• Wilks, Y., "A Preferential Pattern-Seeking Semantics for Natural Language Processing," *Artificial Intelligence,* Vol. 6, 1975, pp. 53-74.

• Winograd, T., *Understanding Natural Language,* New York: Academic Press, 1972.

• Winograd, T., *Language as a Cognitive Process, Vol I: Syntax,* Reading, Mass: Addison-Wesley, 1983.

• Woods, W.A., "Progress in Natural Language Understanding—An Application to Lunar Geology," In *Proc. of the National Computer Conferences,* Montvale, N.J.: AFIPS Press, 1973.

• Woods, W. A., "Cascaded ATN Grammars," *Amer. J. of Computational Linguistics,* Vol. 6, No. 1, 1980, pp. 1-12.

• "Applied Computational Linguistics in Perspective," NRL Workshop at Stanford University, 26-27 June 1981. (Proceedings in *American Journal of Computational Linguistics*, Vol. 8, No. 2, April-June 1982, pp 55-83.)

GLOSSARY

Anaphora: The repetition of a word or phrase at the beginning successive statements, questions, etc.

C.A.I.: Computer-Aided Instruction

Case: A semantically relevant syntactic relationship.

Case Frame: An ordered set of cases for each verb form.

Case Grammar: A form of Transformational Grammar in which the deep structure is based on cases.

Computational Linguistics: The study of processing language with a computer.

Conceptual Dependency (CD): An approach, related to case frames, in which sentences are translated into basic concepts expressed in a small set of semantic primitives.

DB: Data Base

DBMS: Data Base Management System

Deep Structure: The underlying formal canonical syntactic structure, associated with a sentence, that indicates the sense of the verbs and includes subjects and objects that may be implied but are missing from the original sentence.

Discourse: Conversation, or exchange of ideas.

Domain: Subject area of the communication.

Frame: A data structure for grouping information on a whole situation, complex object, or series of events.

Grammar: A scheme for specifying the sentences allowed in a language, indicating the syntactic rules for combining words into well-formed phrases and clauses.

Heuristic: Rule of thumb or empirical knowledge used to help guide a solution.

KB: Knowledge Base

Lexicon: A vocabulary or list of words relating to a particular subject or activity.

Linguistics: The scientific study of language.

Morphology: The arrangement and interrelationship of morphemes in words.

Morpheme: The smallest meaningful unit of a language, whether a word, base or affix.

Network Representation: A data structure consisting of nodes and labeled connecting arcs.

NL: Natural Language

NLI: Natural Language Interface

NLP: Natural Language Processing

NLU: Natural Language Understanding

Parse Tree: A tree-like data structure of a sentence, resulting from syntactic analysis, that shows the grammatical relationships of the words in the sentence.

Parsing: Processing an input sentence to produce a more useful representation.

Phonemes: The fundamental speech sounds of a language.

Phrase Structure Grammar: Also referred to as Context Free Grammar. Type 2 of a series of grammars defined by Chomsky. A relatively natural grammar, it has been one of the most useful in natural-language processing.

Pragmatics: The study of the use of language in context.

Script: A frame-like data structure for representing stereotyped sequences of events to aid in understanding simple stories.

Semantic Grammar: A grammar for a limited domain that, instead of using conventional syntactic constituents such as noun phrases, uses meaningful components appropriate to the domain.

Semantics: The study of meaning.

Sense: Meaning.

Surface Structure: A parse tree obtained by applying syntactic analysis to a sentence.

Syntax: The study of arranging words in phrases and sentences.

Template: A prototype model or structure that can be used for sentence interpretation.

Tense: A form of a verb that relates it to time.

Transformational Grammar: A phrase structure grammar that incorporates transformational rules to obtain the deep structure from the surface structure.